Linear Regression

Regression is the form of predictive modelling where we have to find relation between Independent variables and Dependent variable.

Linear Regression is method to find the best fit line. In Linear Regression dependent variable is continuous in nature.

Technical Definition:

Linear regression is a statistical method used in machine learning to model the relationship between a dependent variable (usually denoted as 'Y') and one or more independent variables (usually denoted as 'X'), by fitting a linear equation to observed data.

If in case, there is only one independent variable to predict dependent variable in such case regression is called as simple linear regression. Whereas if our model based on various independent variables is called Multiple Linear Regression.

The representation of simple linear regression in mathematical equation as

$$y = \alpha + \beta x$$

In this equation:

- y represents the dependent variable.
- x represents the independent variable.
- α is the y-intercept, which is the value of y when x is 0.
- β is the slope of the line, indicating the change in y for a one-unit change in x.

And for multiple linear regression

$$y = \alpha + \beta 1 x1 + \beta 2 x2 + \ldots\ldots + \beta n xn$$

In this equation:

- y represents the dependent variable.
- $x1, x2, ..., xn$ represents the independent variable.
- α is the y-intercept, which is the value of y when x is 0.
- $\beta 1, \beta 2, \ldots., \beta n$ are the slopes of the line, indicating the change in y for a one-unit change in x.


Let's discuss about simple linear regression.

Simple linear regression.

Consider we have dataset represents home prices based on square fit area.

| area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000

1. Our task is to predict the value of house based on above information using linear regression.
2. There can be many line passing through the points but our aim is that to find best fit line which cover all the data points.
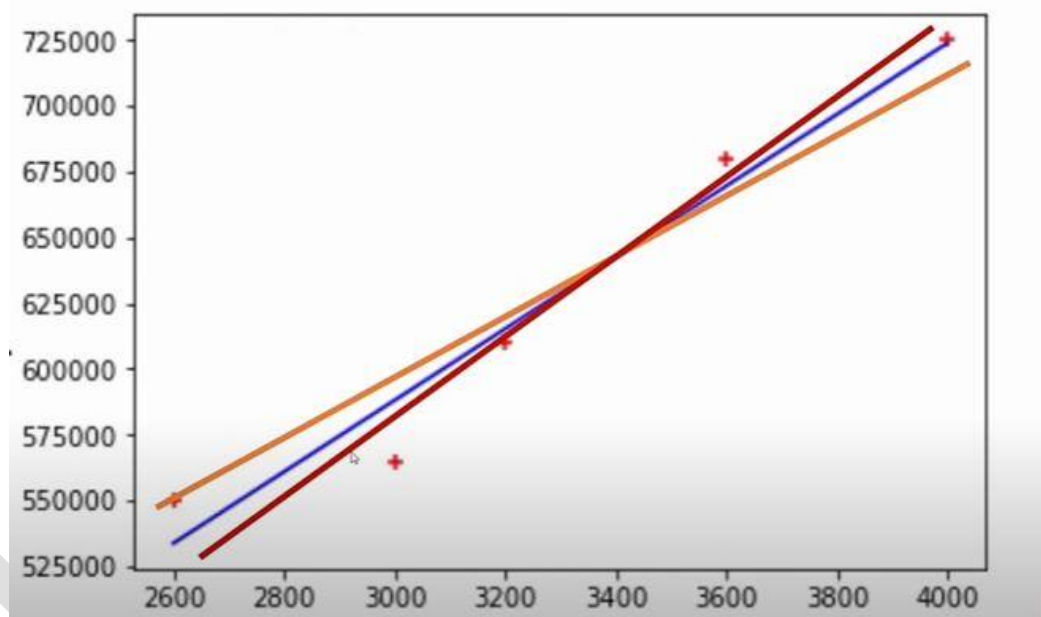
3. Here are the lines passes through our data points, but which is the best choice? Why we have to choose particular one?
4. And here we will get an answer of how linear regression works?
5. Here we will get the value of delta (Δ). where delta is difference between actual value and predicted value by model.
6. We calculate the square of all these delta (Δ) and do sum of all and try to minimize those. For all the lines, we have to follow the same process.

Minimize
$$\sum_{i=1}^{n} (\Delta i)^2$$
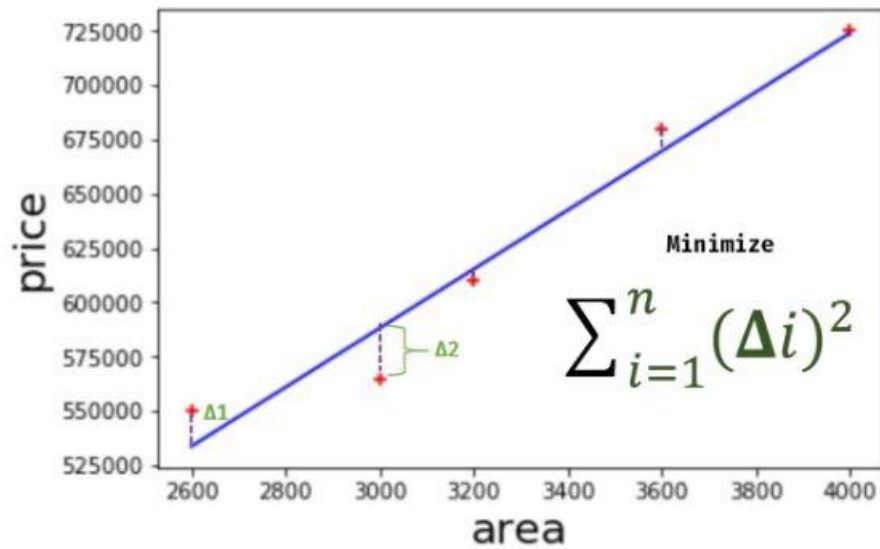
7. After performing this operation on all lines, the line gives us minimum error that is our best fit line.
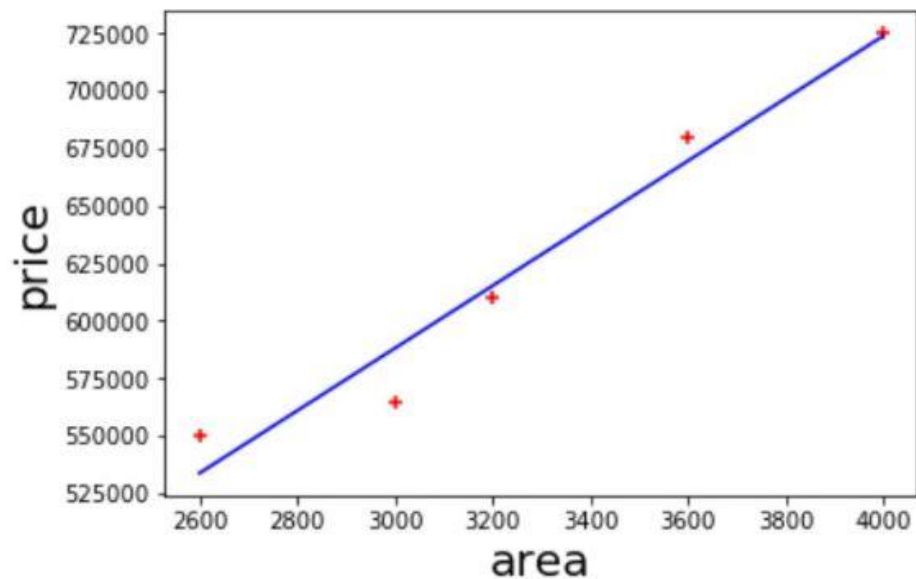
In Simple Terms:

Linear regression is like drawing the best-fit straight line through points on a graph to help us predict things. For example, if we know the size of a house, linear regression helps us guess its price by looking at how other houses' sizes and prices are related.